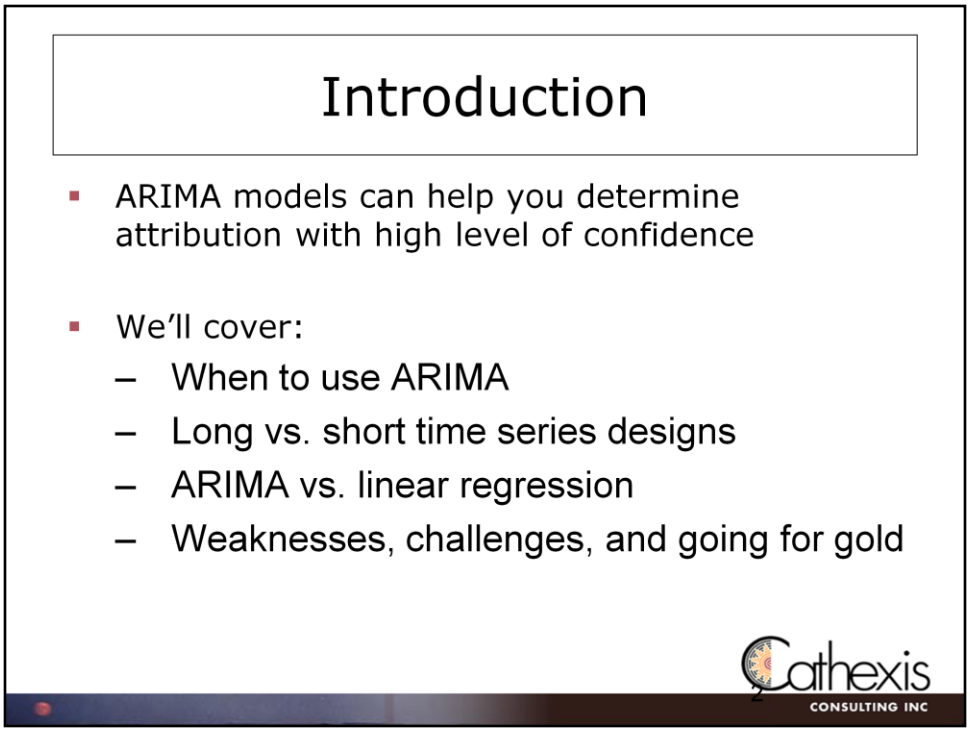
An Introduction to AutoRegressive Integrated Moving Average (ARIMA) Models

By Paul Bakker
paul@cathexisconsulting.ca

Cathexis CONSULTING INC

124 Merton St., Suite 502
Toronto, Ontario M4S 2Z2
Telephone: (416) 469-9954
Fax: (416) 469-8487
E-mail: info@cathexisconsulting.ca

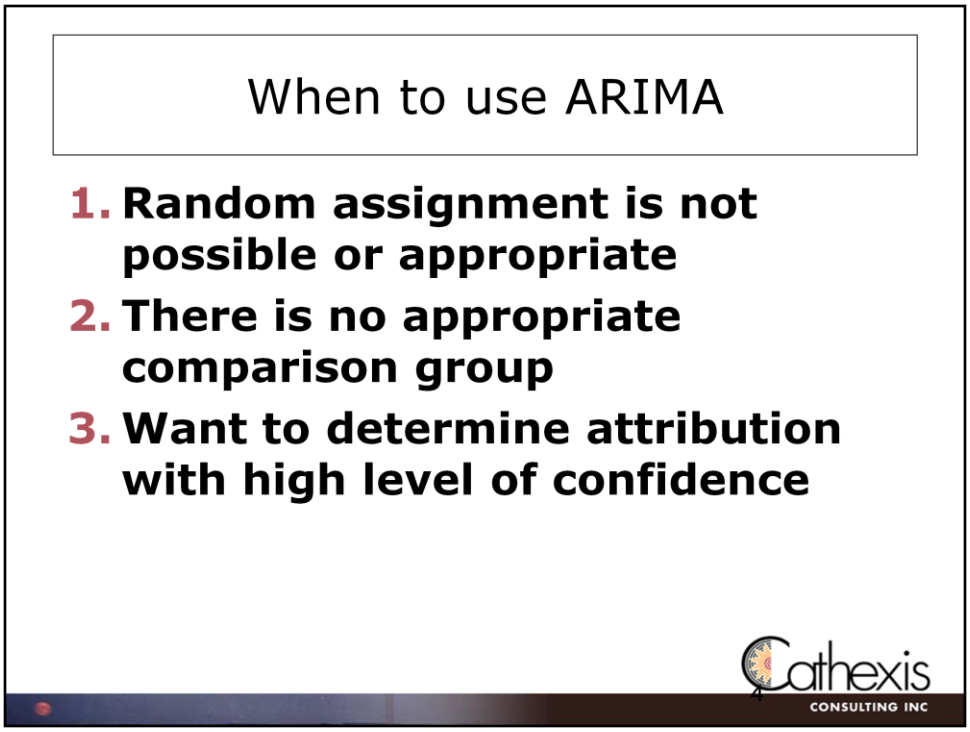


- You've likely tackled the issue of Attribution: whether what you or the programs you are evaluating are creating an impact.
- You likely know this is not an easy task.
- In this presentation I will introduce a method that under particular circumstance can help you determine attribution with a high level of confidence.

While ARIMA models have other uses (i.e. forecasting), this presentation will focus on how the models are used in coordination with interrupted time series designs to discover the effects of interventions.

When to use ARIMA

Cathexis
CONSULTING INC



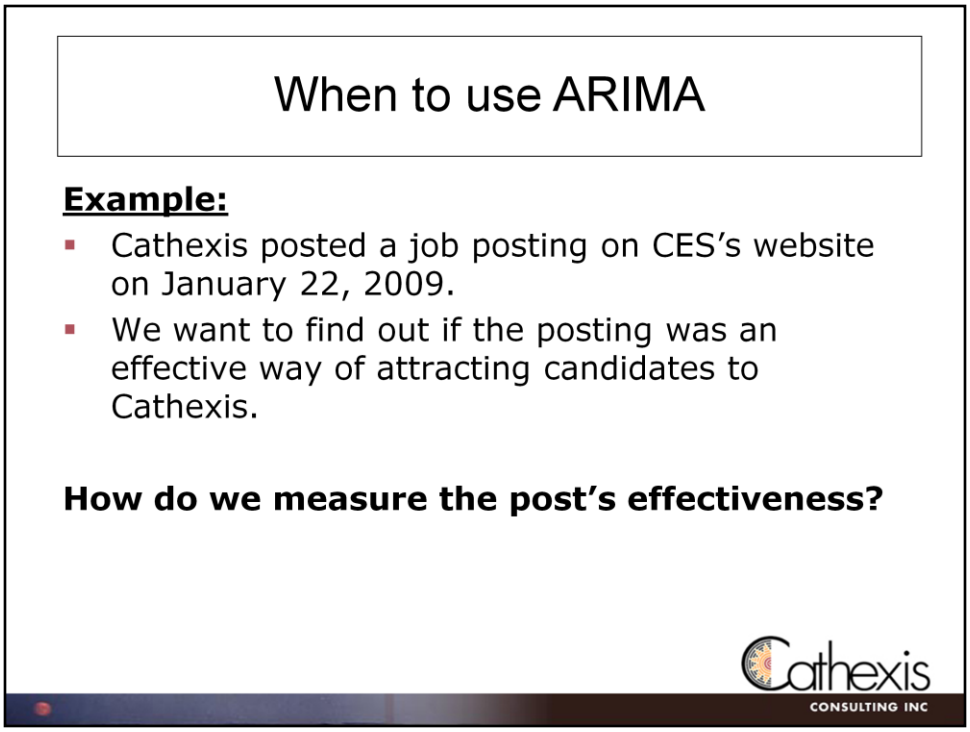
Real world settings often pose methodological, ethical, and practical considerations that make randomized control trials either unfeasible or inappropriate.

- When there is no Appropriate Comparison Group

For example:

- non-treatment groups are known to differ from the treatment groups
- differences between the groups cannot be established.
- Full-coverage programs

•Instead, evaluators can make use of reflexive controls; that is, measures taken prior to an intervention that are used to estimate what would have occurred if the



•Exposure to CES post not random.

•We don't know who did and did not view the post. Can't create a comparison group.

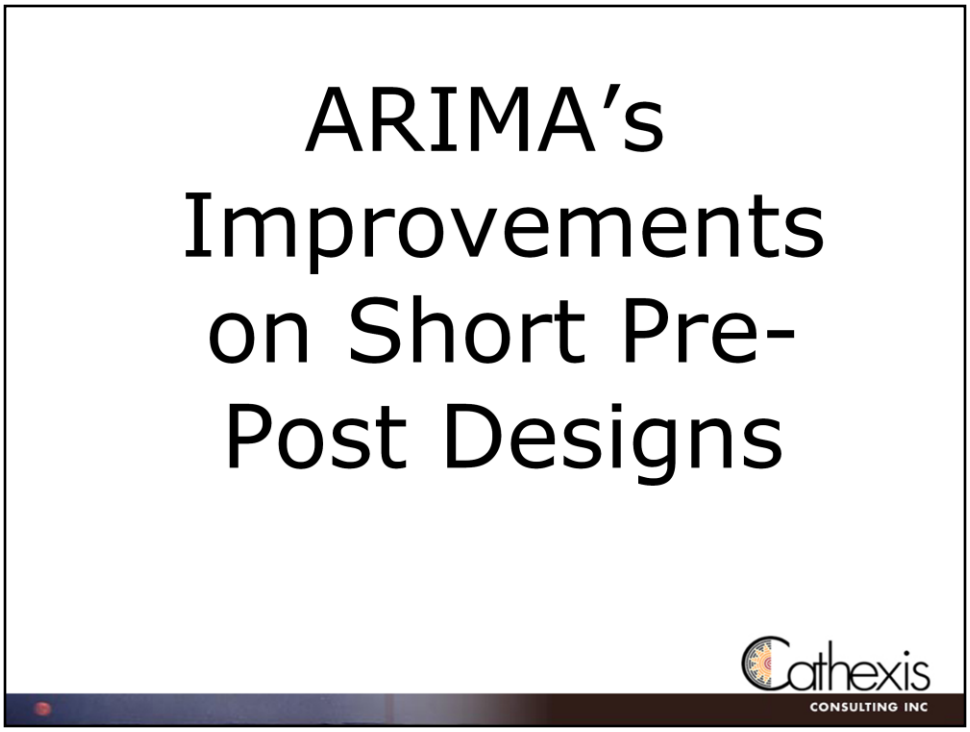
•Can still use a pre-post design.

•We have data on the daily number of visits to Cathexis's website that come from the entire population.

•We can test to see if the website received more visits than normal after the post was posted.

•Example of taking advantage of archival data already collected.

•I'm using this example, because website data has the large #'s of data points required for ARIMA.



I just covered some reasons why you would use a single-group pre-post evaluation design.

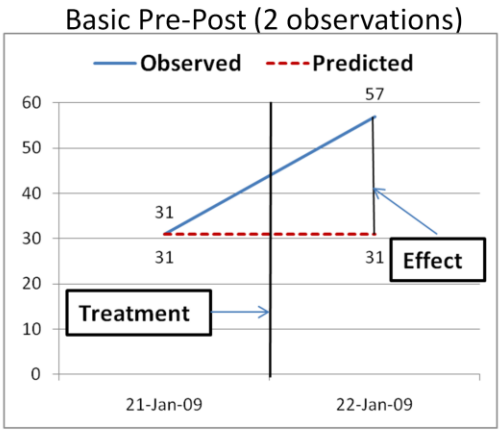
ARIMA is an example such a design; one that uses large numbers of observations.

Next, I'll discuss the benefits of using ARIMA over shorter pre-post designs

ARIMA's Improvements on Short Pre-Post Designs

The Issue of Maturation (a series normal trend)

Were visits normally increasing or decreasing before CES post?



Cathexis CONSULTING INC

There are many reasons why the one group per-test post-test design's conclusions may be wrong.

The main threat addressed by ARIMA modeling is maturation; that is, changes that occur normally as time progresses.

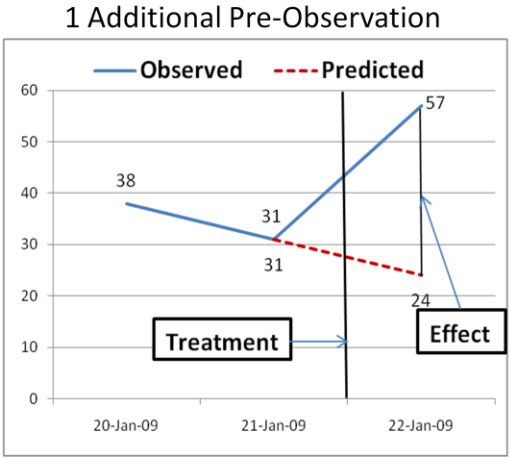
The above design may incorrectly attribute the increase in visits to the CES job posting; whereas, visits to the site could have been undergoing a constantly increasing trend prior to the job posting.

- Assumes that without treatment the post-test score = pre-test score.
- An unlikely assumption.
- Doesn't consider the series's trend.

ARIMA's Improvements on Short Pre-Post Designs

Begin to Consider Trend

But Periodic Changes in Trend May be Normal



Cathexis CONSULTING INC

•Adding more observations before an intervention allow us to begin to consider a series's trend.

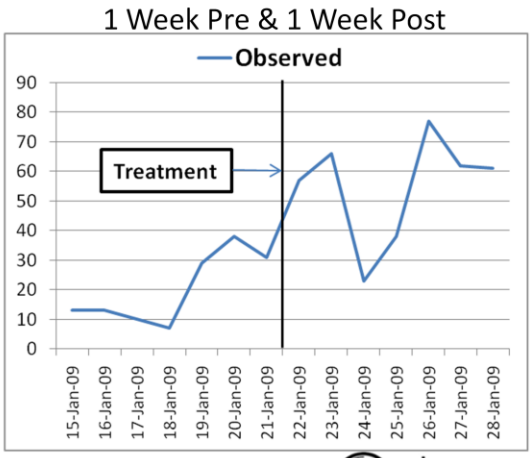
•This design assumes that the rate of decline would continue at the same rate, but that is usually not the case.

•People may be more likely visit the Cathexis website during weekdays. If the CES post was posted on a Sunday, the increase in visits may be normal and not at all due to the post.

ARIMA's Improvements on Short Pre-Post Designs

Clear idea of trend and possible effect

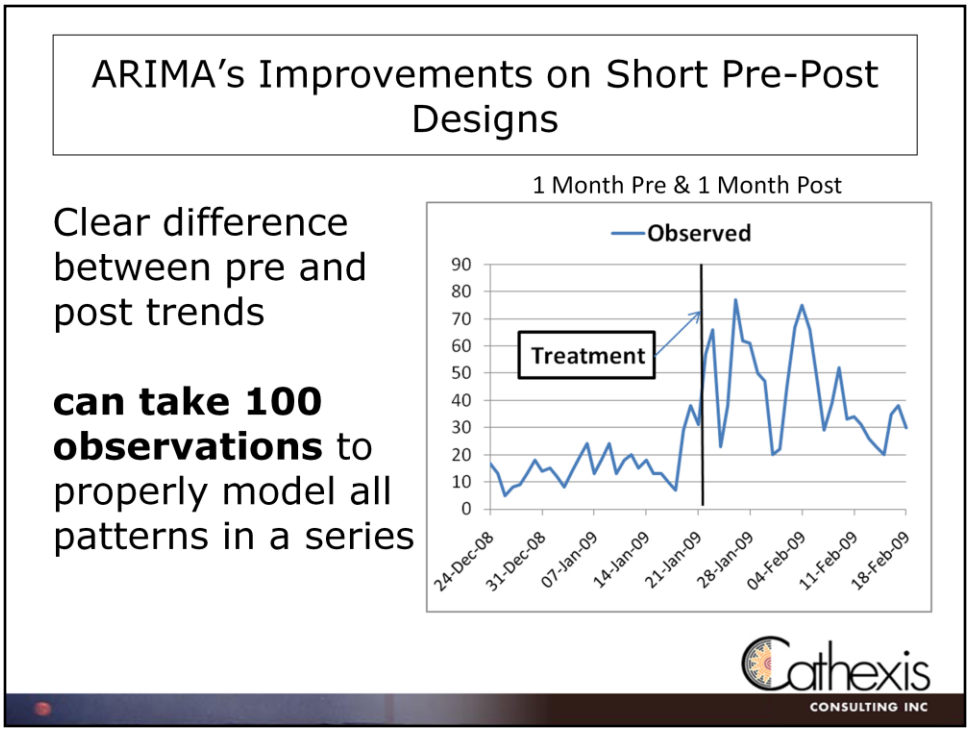
Q's remain: Dip normal? Do pre-post trend continue?



Cathexis CONSULTING INC

Adding more observations gives a clearer idea of the trend, but the trend is still not fully clear.

Trends and effect depend on whether up and down nature of trend is observed in weeks prior and post the observation period.



Adding more observations before an intervention increases the accuracy of one's description of a series's pre-intervention trend; and thus, a more accurate estimate of what would have happened if the intervention did not occur.

Adding more observations after an intervention increases one's ability to determine if a series's trend changed and to describe the nature of that change: whether the slope or level of the series changed, if the change occurred gradually, if there was a delay in the change, and if the change was temporary.

at the very least, need 25 observations before and after an intervention to create an ARIMA model that can accurately describes a series and test whether an intervention made an impact.

ARIMA's Improvements on Linear Regression

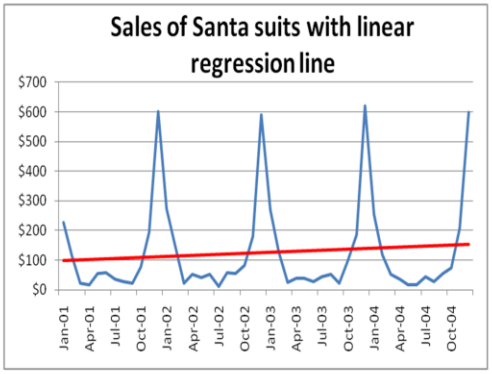
Cathexis CONSULTING INC

ARIMA's Improvements on Linear Regression

The Problem of Autocorrelation

Example: Outliers

- Sales in Dec. > 2 SD from the regression line.
- Considered outliers.
- But normal part of series's trend.



•Autocorrelation or series-correlation refers to the relationships between successive observations. In most point-in-time data information obtained from a unit doesn't depend on the previous observation. Most often in time series data observations made closer together will be more alike than observations made further apart.

•Linear Regression **does not account for** the pattern of **autocorrelation**

•**Leads to errors** in judging when significant changes occur.

•Above is a seasonal pattern of autocorrelation

•Regression considers Dec. sales are outliers: i.e. observations that significantly differ from the normal pattern of variation i.e. 2 SD from regression line.

•ARIMA models discover the patterns in the variation not explained by a regression model & Incorporate those patterns into its model.

•With accurate description of all the patterns in the data, ARIMA models can conduct accurate significance tests.

- I`ve talked about how using a long time series controls accounts for a series normal trend and ARIMA models are better able than linear regression to model that trend.
- Next we`ll discuss the possible weaknesses and challenges of ARIMA designs and how they can be overcome.

History: change caused by something else that happened at around same time as intervention

Limit by decreasing time between observations, intervention occurs abruptly not gradually, and no delay in effect.

Instrumentation: method of measurement changes around same time as intervention.

Limit by ensuring measurement methods do not change.

Selection: characteristics of target population change around the same time as intervention.

Test to see if group characteristics change when intervention introduced.

We can limit threats by using strong theory and measuring extraneous changes. Possible to create design so that threats are vey unlikely.

For instance, our example of the CES job posting.

History: we haven`t measured every other event that occurred that day, so we are using some theory that the change is due to our Job posting. I would argue that our theory is quite strong.

Need to wait at least 25 years to use ARIMA if working with yearly data.

Is possible to get more frequent data: i.e. website data.

another example: Early Return to Work.

keeping track of when each worker injured and when returned to work.

Create database that can group injured workers on a short time-period (perhaps weekly).

Need to consider making treatment mandatory and enforced.

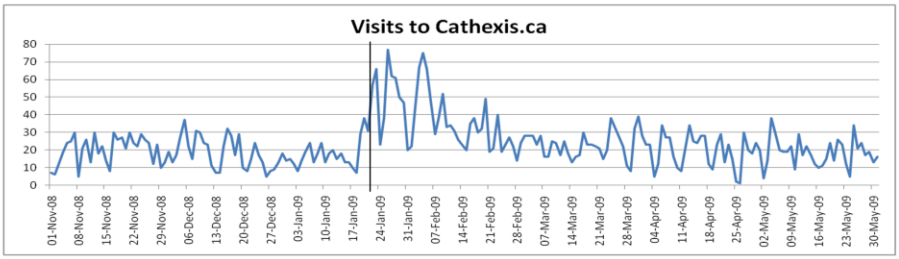
If optional and slowly adopted, it will be more difficult to determine impact of treatment vs. impact of other changes occurring during uptake.

Relatively recently some stats programs began automating ARIMA analysis. (SPSS starting with 14, SAS, maybe others).

ARIMA Analysis Using SPSS's Expert Modeler.

Back to our website example

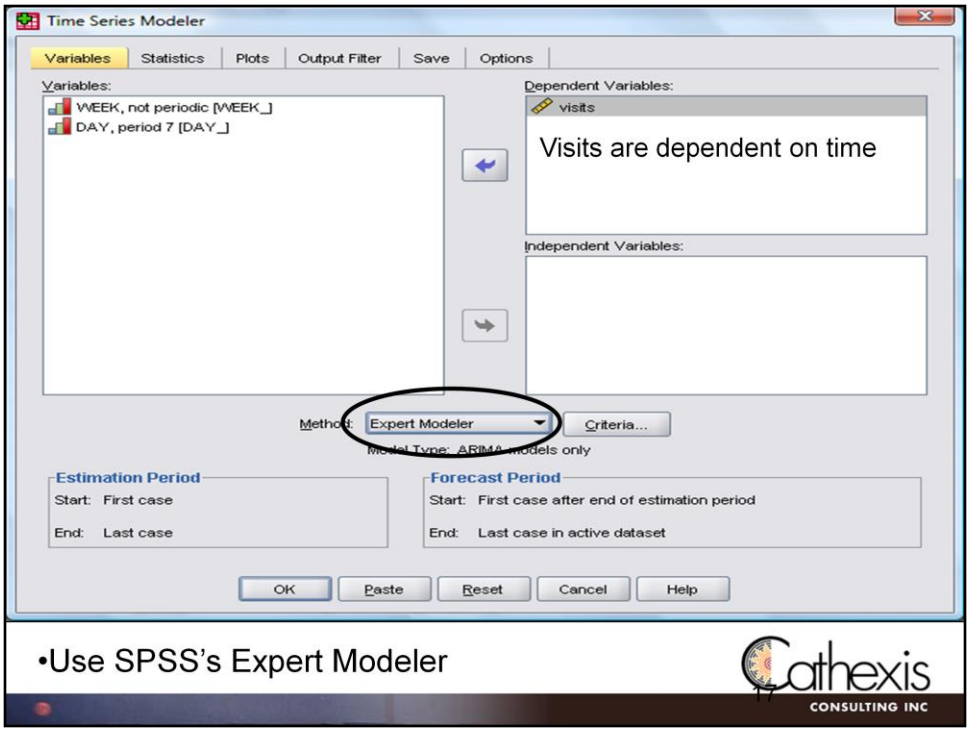
An ARIMA model will allows to know when and by how much visits increased with given level (95%) of confidence.



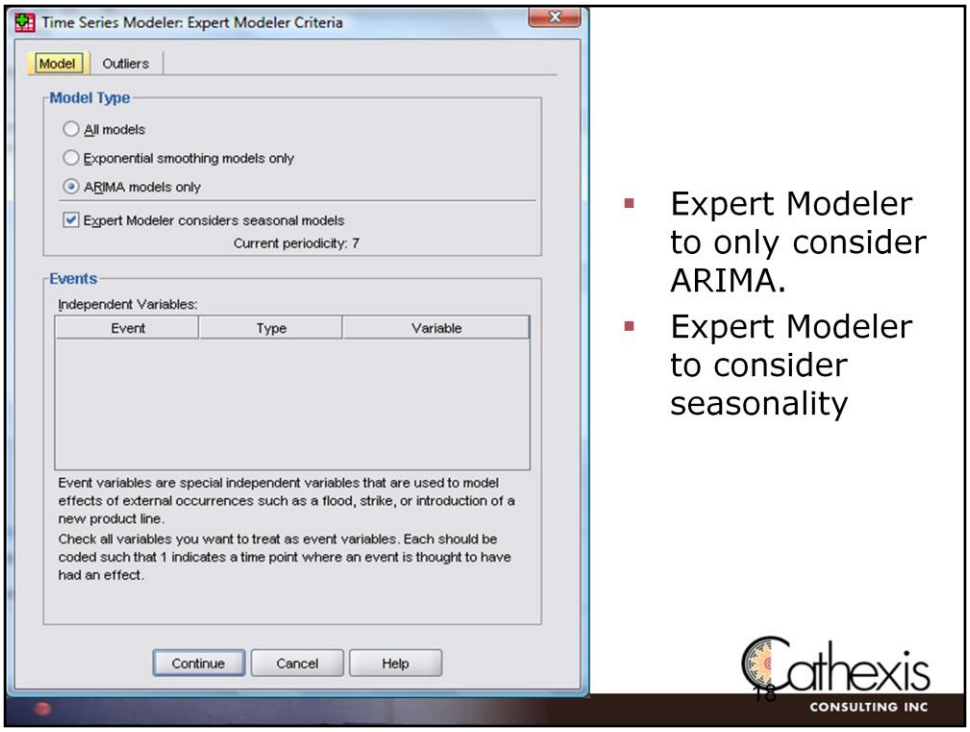
Back to our website example.

daily data on the # of visits to Cathexis.ca from Nov. 1st to May 31st, 2009.

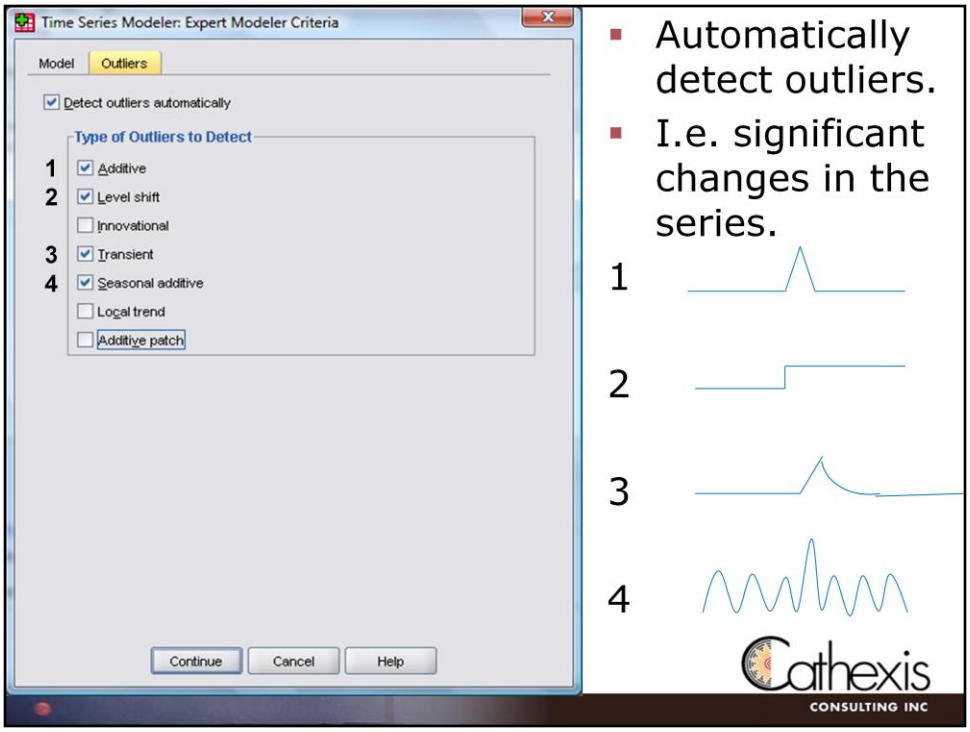
- It is easy to visually notice the change in this series, but:
  - Did something occur a day or two before the CES post to drive more visits?
  - What is the extent of the change?



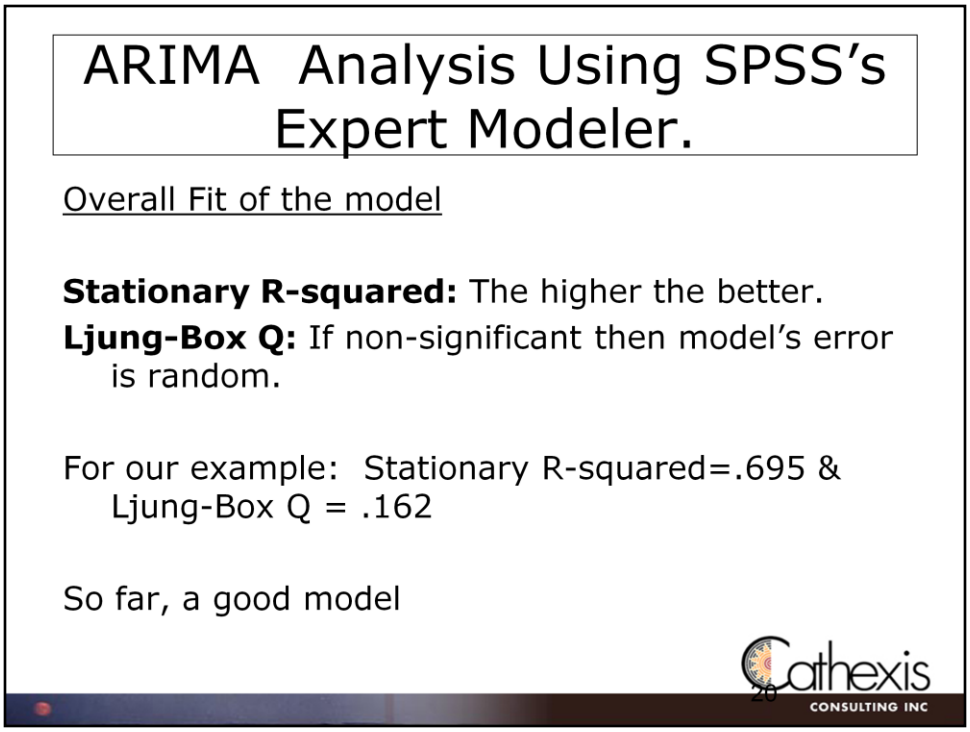
Select **Analyze – Forecasting – Create Models** to find the best fitting ARIMA model for the data.



We have weekly data.



Again, outliers are significant changes in a series.



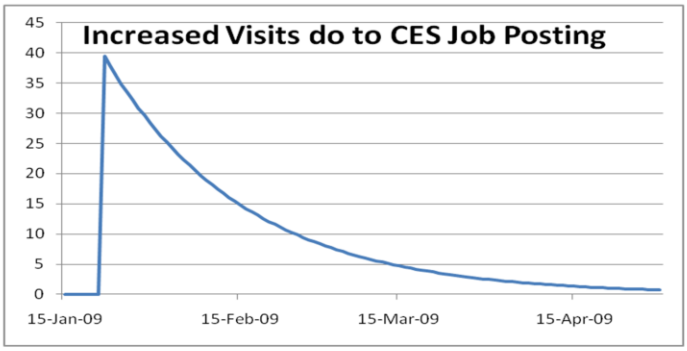
I'm only going to go over the important elements to judge appropriateness of the model and to understand the nature of the changes to the series.

Stationary R-squared: how well the model predicts observed values.

Ljung-Box Q: test of whether error not explained by the model is random.

ARIMA Analysis Using SPSS's Expert Modeler.

The model determines Jan. 22 (day of CES post) to be a transient outlier with a magnitude of 39.4 with a decay factor of 96%



24

Cathexis CONSULTING INC

Which means on the 22nd 39 more visits than expected occurred, on the 23rd 38 (39*.96) more visits than expected occurred, and on the 24th 36 (38*.96) occurred, and so on

So the increase didn't occur prior to CES post, and we know how much the post resulted in visits to Cathexis's website.

Summary

- We'll covered:
  - When to use ARIMA
  - Long vs. short time series designs
  - ARIMA vs. linear regression
  - Weaknesses, challenges, and going for gold

- ARIMA models make it is possible to make claims of attribution with a high level of confidence without random assignment or comparison groups

Cathexis CONSULTING INC